

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-271392

(43)Date of publication of application : 20.10.1995

(51)Int.Cl. G10L 3/00  
G10L 3/00

(21)Application number : 06-058515

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>

(22)Date of filing : 29.03.1994

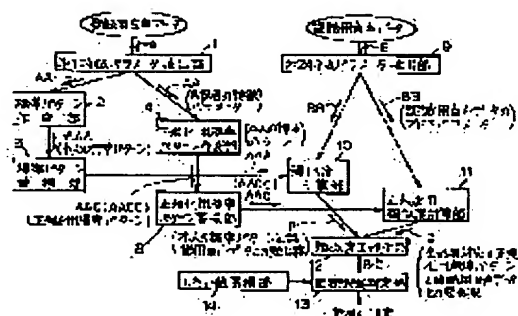
(72)Inventor : MATSUI TOMOKO  
FURUI SADAHIRO

## (54) DEGREE OF SIMILARITY NORMALIZATION METHOD FOR SPEAKER RECOGNITION AND SPEAKER RECOGNITION DEVICE USING THE METHOD

### (57)Abstract:

PURPOSE: To provide a speaker recognition degree of similarity normalization method and a speaker recognition device employing the above method in which the amount of computations is reduced and the recognition rate is improved.

CONSTITUTION: Standard patterns corresponding to all registered speakers are generated as normalized standard patterns AAC or AACC. The degree of similarity of recognition voice data B with the standard patterns for normalization is computed as the normalization degree of similarity Q. Based on a posteriori probability concept, a standard pattern AAA of each speaker and a degree of similarity P of the recognition voice data are divided by the degree of similarity P of the recognition voice data ( $P/Q$ ) and the dispersion of the degree of similarity between the standard pattern of each speaker and the voice data for recognition is normalized.



### LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-271392

(43) 公開日 平成7年(1995)10月20日

(51) Int.Cl.<sup>6</sup>

G 1 0 L 3/00

識別記号

5 3 1 F

5 2 1 G

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数 4 O L (全 6 頁)

(21) 出願番号 特願平6-58515

(22) 出願日 平成6年(1994)3月29日

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72) 発明者 松井 知子

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72) 発明者 古井 貞熙

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(74) 代理人 弁理士 草野 卓

(54) 【発明の名称】 話者認識用類似度正規化方法及びこの方法を用いた話者認識装置

(57) 【要約】

【目的】 計算量を少なくすることができ、認識率を高めることができる話者認識用類似度正規化方法及びこの方法を用いた話者認識装置を提供する。

【構成】 全登録話者対応の標準パターンを正規化用標準パターンAACまたはAACCCとして作成し、認識用音声データBを正規化用標準パターンとの類似度を正規化用類似度Qとして算出し、事後確率の考えに基づき、各話者の標準パターンAAAと認識用音声データの類似度Pを正規化用類似度で除す ( $P/Q$ ) ことにより、各話者の標準パターンと認識用音声データとの類似度のばらつきを正規化する。

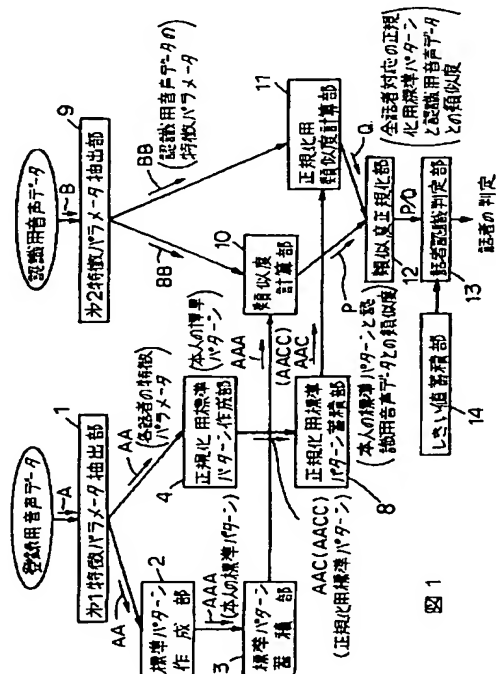


図 1

## 【特許請求の範囲】

【請求項 1】 認識用音声データを特徴パラメータを用いた表現形式に変換し、その表現形式による認識用音声データと、あらかじめ話者対応に登録された上記表現形式による特徴パラメータの標準パターンとの類似度を求めて、上記認識用音声データを発声した話者を認識する話者認識方法において、

あらかじめ全登録話者対応の標準パターンを正規化用標準パターンとして作成し、上記認識用音声データと上記正規化用標準パターンとの類似度を正規化用類似度として算出し、事後確率の考えに基づき、各話者の標準パターンと上記認識用音声データの類似度を上記正規化用類似度で除することにより、各話者の標準パターンと上記認識用音声データとの類似度のばらつきを正規化することを特徴とする話者認識用類似度正規化方法。

【請求項 2】 A. 登録用音声データを特徴パラメータを用いた表現形式の音声データに変換する第 1 特徴パラメータ抽出部と、

B. 上記表現形式による特徴パラメータの各話者に対応した標準パターンを作成する標準パターン作成部と、

C. この標準パターン作成部で作成した各話者に対応した標準パターンを蓄積する標準パターン蓄積部と、

D. 上記第 1 特徴パラメータ抽出部で時系列に変換した各話者の特徴パラメータを用いた表現形式の登録用音声データから全登録話者用標準パターンを作成する正規化用標準パターン作成部と、

E. この正規化標準パターン作成部で作成した正規化用標準パターンを蓄積する正規化用標準パターン蓄積部と、

F. 認識用音声データを特徴パラメータを用いた表現形式の音声データに変換する第 2 特徴パラメータ抽出部と、

G. この第 2 特徴パラメータ抽出部で変換した特徴パラメータを用いた表現形式の認識用音声データと上記正規化用標準パターン蓄積部に蓄積した全話者対応の正規化用標準パターンとの類似度を算出する正規化用類似度計算部と、

H. 上記標準パターン蓄積部に蓄積した各話者対応の標準パターンと、上記第 2 特徴パラメータ抽出部で変換した認識用音声データの特徴パラメータを用いた表現形式の認識用音声データとの類似度を算出する類似度計算部と、

I. この類似度計算部で算出した類似度を上記正規化類似度計算部で算出した正規化用類似度で除算する類似度正規化部と、

J. この類似度正規化部で正規化された類似度と、しきい値蓄積部に蓄積した各話者対応のしきい値と比較し、話者を特定する話者認識判定部と、

によって構成したことを特徴とする話者認識装置。

【請求項 3】 請求項 2 記載の正規化用標準パターン作

成部を、全登録話者用標準パターン作成部によって構成したことを特徴とする話者認識装置。

【請求項 4】 請求項 2 記載の正規化用標準パターン作成部を、標準パターン話者適応部と標準パターン平均部とによって構成したことを特徴とする話者認識装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】この発明は、例えばインターホンの音声から訪問者が誰であるかを認識したり、入力された音声により暗証番号の人と同一人であることを同定したりするためなどに用いられ、認識用音声データを、特徴パラメータを用いた表現形式に変換し、その表現形式による認識用音声データと、あらかじめ話者対応に登録された上記表現形式による標準パターンとの類似度を求めて入力音声を発声した話者を認識する話者認識方法における類似度の正規化方法及びこの正規化方法を用いた話者認識装置に関する。

## 【0002】

【従来の技術】話者認識は、話者が発声した文章などの音声に含まれる特徴パラメータ（例えばケプストラム、ピッチなど）を求める。登録話者の特徴パラメータの標準パターンとの類似度によって判定する手法がよく用いられる。この類似度は、発声内容、収録時期、伝送系、マイクロホンなどの違いによって大きく変動するために、話者認識性能を低下させてきた。

【0003】従来、この類似度の正規化のために、認識用音声をも本人を含めた複数の話者の標準パターンと比較して類似度を計算し、その上位  $n$  名（ $n$  は 1 以上の整数）の平均類似度で、各話者の標準パターンとの類似度を除することによって類似度のばらつきを正規化する方法（例えば文献「松井知子、古井貞照、「連結音韻モデルによる話者認識」、信学技報、SP92-128, 1993」）などが試みられてきた。

## 【0004】

【発明が解決しようとする課題】しかしながら、この従来の方法では、全ての登録話者それぞれの標準パターンとの類似度を計算する必要があり、正規化のための計算量が登録話者数に比例して大きくなるという問題があった。

## 【0005】

【課題を解決するための手段】認識用音声データを特徴パラメータを用いた表現形式に変換し、その表現形式による認識用音声データと、あらかじめ話者対応に登録された上記表現形式による特徴パラメータの標準パターンとの類似度を求めて、上記認識用音声データを発声した話者を認識する話者認識方法において、あらかじめ全登録話者対応の標準パターンを正規化用標準パターンとして作成し、認識用音声データと正規化用標準パターンとの類似度を正規化用類似度として算出し、事後確率の考え（※）に基づき、各話者の標準パターンと上記認識用

3

音声データの類似度を上記正規化用類似度で除すことにより、各話者の標準パターンと上記認識用音声データとの類似度のばらつきを正規化することを特徴とする話者認識用類似度正規化方法を提案する。

#### 【0006】※事後確率の考え

本来、入力音声とある話者の標準パターンとの類似度は、例えば文献「竹内啓編集、“統計学辞典”、東洋経済新報社 pp.13-17, 1989」で述べられている確率論的には、入力音声  $x$  が与えられた時、それがその話者  $s_c$  の声である確率  $P(s_c | x)$  として計算されるべきだが、実際にはある話者の標準パターンが与えられた時の入力音声の確率  $P(x | s_c)$  しか観測することができず、一般的には後者の確率  $P(x | s_c)$  を使って話者の判定を行う。しかし、後者の確率  $P(x | s_c)$  は、発声内容、収録時期などの違いによってばらつく。事後確率の考えに基づく正規化では、前者の確率  $P(s_c | x)$  を

$$P(s_c | x) = P(x | s_c) \times P(s_c) / \sum_i (P(x | s_i) \times P(s_i))$$

$s_i$  : 任意の話者

のように展開し、観測可能な後者の確率  $P(x | s_c)$  を  $\sum_i (P(x | s_i) \times P(s_i))$  で除し、前者の確率  $P(s_c | x)$  を求める。 $P(s_i)$  は、話者  $s_i$  の出現確率である。ただし、 $\sum_i (P(x | s_i) \times P(s_i))$  を求めるためには、未登録の話者も含め、その話者認識システムを利用する全話者について確率の和を計算する必要がある。従来方法（例えば文献「松井知子、古井貞照、“連結音韻モデルによる話者認識”、信学技報、SP92-128, 1993」）では、計算量を削減するために、話者の出現確率  $P(s_i)$  は各話者共通であるとして、分母と分子でキャンセルし、入力音声を含めた全話者の標準パターンと比較して類似度を計算し、 $\sum_i P(x | s_i)$  をその上位  $n$  名 ( $n$  は 1 以上の整数) の平均類似度で近似していた。この発明では、 $\sum_i P(x | s_i)$  を入力音声と全登録話者用の標準パターンとの類似度で近似する。

【0007】更にこの発明では、登録用音声データの特徴パラメータを用いた表現形式の音声データに変換する第 1 特徴パラメータ抽出部と、表現形式による特徴パラメータの各話者に対応した標準パラメータを作成する標準パターン作成部と、この標準パターン作成部で作成した各話者に対応した標準パターンを蓄積する標準パターン蓄積部と、第 1 特徴パラメータ抽出部で時系列に変換した各話者の特徴パラメータを用いた表現形式の登録用音声データから全登録話者用標準パターンを作成する正規化用標準パターン作成部と、この正規化標準パターン作成部で作成した正規化用標準パターンを蓄積する正規化用標準パターン蓄積部と、認識用音声データの特徴パラメータを用いた表現形式の音声データに変換する第 2 特徴パラメータ抽出部と、この第 2 特徴パラメータ抽出

4

部で変換した特徴パラメータを用いた表現形式の認識用音声データと上記正規化用標準パターン蓄積部に蓄積した全話者対応の正規化用標準パターンとの類似度を算出する正規化用類似度計算部と、標準パターン蓄積部に蓄積した各話者対応の標準パターンと、第 2 特徴パラメータ抽出部で変換した認識用音声データの特徴パラメータを用いた表現形式の認識用音声データとの類似度を算出する類似度計算部と、この類似度計算部で算出した類似度を正規化類似度計算部で算出した正規化用類似度で除算する類似度正規化部と、この類似度正規化部で正規化された類似度と、しきい値蓄積部に蓄積した各話者対応のしきい値と比較し、話者を特定する話者認識判定部と、によって構成した話者認識装置を提案する。

#### 【0008】

【作用】この発明による正規化方法及び装置によれば、あらかじめ全登録話者対応の正規化用標準パターンを作成し、認識用音声データと正規化用標準パターンとの類似度を正規化用類似度として求め、各話者ごとの標準パターンと認識用音声データとの類似度を、事後確率の考えに基づいて正規化用類似度で除すことにより、各話者ごとの類似度のばらつきを正規化するので、正規化のための類似度の計算は 1 回で済む。

【0009】また、この発明によれば、発声内容が任意のテキスト独立型（例えば文献「松井知子、古井貞照：“VQ、離散／連続 HMM によるテキスト独立形話者認識法の比較検討”、電子情報通信学会音声研究会資料、SP91-89, 1991」）、認識のたびに発声内容が指示されるテキスト指定型（例えば文献「松井知子、古井貞照、“連結音韻モデルによる話者認識”、信学技報、SP92-128, 1993」）など、種々の話者認識方法に適用することができる。

#### 【0010】

【実施例】図 1 を用いてこの発明の一実施例を説明する。図 1 に示す実施例ではテキスト独立型話者認識を例示して示す。この発明では、図 1 に示すように各話者ごとに、登録用音声データ A を第 1 特徴パラメータ抽出部 1 に入力する。第 1 特徴パラメータ抽出部 1 では、入力された登録用音声データ A を例えばケプストラム、ピッチなどの特徴パラメータを用いた表現形式に変換する。次に、特徴パラメータの時系列に変換された登録用音声データ A が、標準パターン作成部 2 に入力され、登録用音声データ A に含まれた特徴パラメータの標準パターンが、例えば複数のガウス分布の重み付き加算などで表現される。

【0011】複数のガウス分布の重み付き加算を作成する方法として、例えば文献「松井知子、古井貞照：“VQ、離散／連続 HMM によるテキスト独立形話者認識法の比較検討”、電子情報通信学会音声研究会資料、SP91-89, 1991」に述べられている方法などを用いることができる。各話者の標準パターン A A A を話者別に標準パタ

ーン蓄積部 3 に蓄える。

【0012】並行して、第 1 特徴パラメータ抽出部 1 で特徴パラメータの時系列に変換された各話者の登録用音声データ A A を、正規化用標準パターン作成部 4 に入力する。正規化用標準パターン作成部 4 では、例えば図 2 または図 3 に示す方法で、全話者登録話者対応の正規化用標準パターン A A C (又は A A C C) を作成する。図 2 に示す例は、全登録話者用標準パターン作成部 5 で作成される全話者登録話者対応の正規化用標準パターン A A C は、全話者の登録用音声データ A の特徴パラメータ

の分布を表す、例えば複数のガウス分布の重み付き加算として作成される。その具体的な方法としては、例えば上記文献に述べられている方法などを用いることができる。

【0013】また図 3 に示す例では、あらかじめ多数話者の音声から作成した、例えば複数のガウス分布の重み付き加算として表現された一般的な標準パターンを用いて、各話者ごとに、標準パターン話者適応部 6 で、その登録用音声データ A に合わせて、例えばそれらのガウス分布の重みの値を変更する。その具体的な方法として

は、例えば文献「X. D. Huang: "Phoneme Classification Using Semi-continuous Hidden Markov Models," IEEE Trans. ASSP. Vol. 40, No. 5, pp. 1062-1067, 1992」に述べられている方法などを用いることができる。標準パターン平均部 7 では、各話者ごとに変更した部分、例えばガウス分布の重みの値について、全話者で平均をとる。このようにして得られた全登録話者対応の正規化用標準パターン A A C C は、その平均した標準パターン、例えばその平均重みによる複数のガウス分布の重み付き加算として表現される。

【0014】図 2 の方法では、新しい話者を登録するたびに、全登録話者対応の正規化用標準パターン A A C を一から作成しなおす必要があるが、図 3 の方法では、その話者についてだけ、標準パターン話者適応部 6 で、一般的な標準パターンをその登録用音声に合わせて変更し、標準パターン平均部 7 で、すでに計算されている他の話者の標準パターンと併せて、その平均を計算するだけでよい。全登録話者対応の正規化用標準パターン A A C または A A C C は正規化用標準パターン蓄積部 8 に蓄える。

【0015】次に、話者を認識する段階では、認識用音声データ B を特徴パラメータ抽出部 9 に入力する。特徴パラメータ抽出部 9 では、入力された認識用音声データ B を特徴パラメータ抽出部 1 と同じ表現形式に変換する。特徴パラメータ抽出部 9 で得られた特徴パラメータ B B の時系列は、類似度計算部 10 と正規化用類似度計算部 11 に入力される。

【0016】類似度計算部 10 では、標準パターン蓄積部 3 に蓄えられた本人の標準パターン A A A と、認識用音声データ B を変換した特徴パラメータ B B との類似の

度合いを計算する。正規化用類似度計算部 11 では、正規化用標準パターン蓄積部 8 に蓄えられた全登録話者対応の正規化用標準パターン A A C または A A C C と、認識用音声データ B を変換した特徴パラメータ B B との類似の度合い P と Q を計算する。

【0017】この具体的方法としては、例えば文献「松井知子、古井貞熙: "音源・声道特徴を用いたテキスト独立形話者認識"、電子情報通信学会音声研究会資料、SP90-26, 1990」に述べられている方法などを用いることができる。それぞれ計算された類似度の値 P と Q は、類似度正規化部 12 に入力される。類似度正規化部 12 では、本人の標準パターン A A A に対する類似度の値 P を、全登録話者用の標準パターン A A C (又は A A C C) に対する類似度の値 Q で除することにより、本人の標準パターンに対する類似度の値 P を正規化する。その正規化された類似度の値  $P/Q$  は話者認識判定部 13 に送られ、話者の判定を行う。話者認識判定部 13 では、しきい値蓄積部 14 から、その本人の声とみなせる類似度の変動の範囲を示すしきい値を読み出して、類似度の値  $P/Q$  と比較し、その類似度の値  $P/Q$  が読み出されたしきい値よりも大きければ本人の音声であると判定し、しきい値よりも小さければ他人の音声であると判定する。

【0018】次に実験例を述べる。実験は、男性 20 名、女性 10 名が約 10 ヶ月に渡る 5 つの時期 (時期 A, B, C, D, E) に発声した文章データ (1 文章長は平均 4 秒) を対象とする。登録話者として男性 10 名、女性 5 名、詐称者としてその他の男性 10 名、女性 5 名を用いた。これらの音声を従来使われている特徴量、つまりケプストラムの細かい時間ごとの時系列に変換する。ケプストラムは標準化周波数 12 kHz、フレーム長 32 ms、フレーム周期 8 ms、LPC 分析 (Linear Predictive Coding, 線形予測分析) 次数 16 で抽出した。学習には、時期 A に発声した 10 文章を用い、テストでは、時期 B, C, D, E に発声した 5 文章を 1 文章ずつ用いた。

【0019】この発明の効果は、テキスト独立型 (例えば文献「松井知子、古井貞熙: "VQ, 離散/連続 HMM によるテキスト独立形話者認識法の比較検討"、電子情報通信学会音声研究会資料、SP91-89, 1991」)、テキスト指定型 (例えば文献「松井知子、古井貞熙: "連結音韻モデルによる話者認識"、信学技報、SP92-128, 1993」) の話者認識において試した。テキスト独立型話者認識では、各話者の標準パターンは 64 個のガウス分布の重み付き加算 (例えば文献「松井知子、古井貞熙: "VQ, 離散/連続 HMM によるテキスト独立形話者認識法の比較検討"、電子情報通信学会音声研究会資料、SP91-89, 1991」) で表した。テキスト指定型話者認識では、各話者の標準パターンは、各音素ごとに 3 状態、256 個のガウス分布の重み付き加算 (例えば文献「松井

知子、古井貞熙、“連結音韻モデルによる話者認識”、信学技報、SP92-128、1993」)で表した。

【0020】結果は、テキスト独立型話者認識では平均話者照合誤り率で、テキスト指定型話者認識では平均話者・テキスト照合誤り率で評価した。その結果を図3に示す。これより、この発明方法は類似度の正規を施さない場合と比較して、平均照合誤り率がほぼ半分になった。また、文献「松井知子、古井貞熙、“連結音韻モデルによる話者認識”、信学技報、SP92-128、1993」に述べられている従来法と比べ、平均照合誤り率は等しいか、あるいはそれ以下であり、しかも正規化に必要とされる計算量は本方法の方が圧倒的に少ない。これらの結果より、発明方法は有効であることが実証された。

【0021】

【発明の効果】以上述べたように、この発明においては全話者登録話者対応の正規化用標準パターンとの類似度を使って、本人の標準パターンに対する類似度の値を1回の計算で正規化しており、計算量を少なくすることができる。また発声内容、収録時期、伝送系、マイクロホンなどの違いによる類似度の変動の影響を受け難い話者認識を行うことができる。

【図面の簡単な説明】

【図1】この発明の一実施例を示すブロック図。

【図2】図1に示した実施例に用いた正規化用標準パターン作成部における標準パターン作成方法を説明するためのブロック図。

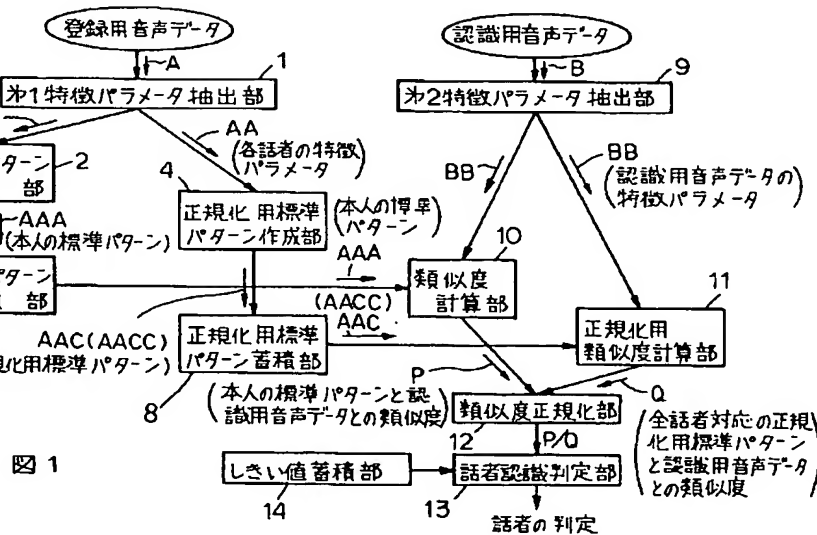
【図3】図2と同様のブロック図。

【図4】この発明の効果を説明するための実験結果を示す図である。

【符号の説明】

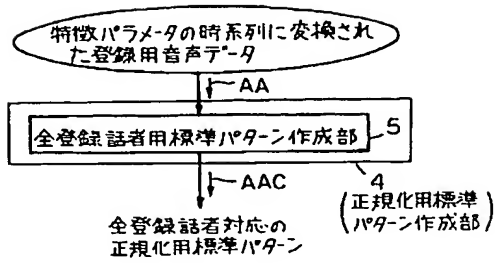
- A 登録用音声データ
- AA 各話者の特徴パラメータ
- AAA 本人の標準パターン
- AAC, AACCC 正規化用標準パターン
- B 認識用音声データ
- BB 認識用音声データの特徴パラメータ
- P 本人の標準パターンと認識用音声データとの類似度
- Q 全話者対応の正規化用標準パターンと認識音声データとの類似度
- 1 第1特徴パラメータ抽出部
- 2 標準パターン作成部
- 3 標準パターン蓄積部
- 4 正規化用標準パターン作成部
- 5 全登録話者用標準パターン作成部
- 6 標準パターン話者適応部
- 7 標準パターン平均部
- 8 正規化用標準パターン蓄積部
- 9 第2特徴パラメータ抽出部
- 10 類似度計算部
- 11 正規化用類似度計算部
- 12 類似度正規化部
- 13 話者認識判定部
- 14 しきい値蓄積部

【図1】



【図 2】

図 2



【図 4】

図 4：話者（・テキスト）照合の平均誤り率（％）

時期	テキスト独立型			テキスト指定型		
	正規化なし	従来法	本方法	正規化なし	従来法	本方法
B	3.0	0.6	0.8	1.8	1.8	0.8
C	4.1	1.3	1.7	4.0	1.5	1.2
D	4.9	1.9	2.0	3.7	2.2	1.1
E	5.0	2.7	1.9	2.1	3.0	1.4
平均	4.3	1.6	1.6	2.9	2.1	1.1

【図 3】

図 3

